

## CHILDREN'S PRODUCTION OF DETERMINERS AS A TEST CASE FOR INNATE SYNTACTIC CATEGORIES

CATRIONA SILVEY

*Department of Psychology, University of Chicago  
Chicago, USA  
casilvey@uchicago.edu*

CHRISTOS CHRISTODOULOPOULOS

*Department of Computer Science, University of Illinois at Urbana-Champaign  
Urbana, USA  
christod@illinois.edu*

A central debate in language evolution is whether humans have a specific innate capacity for language, or whether domain-general learning abilities can explain the acquisition of linguistic structures. One testing ground for these hypotheses has been children's early use of English determiners, specifically the definite and indefinite articles 'the' and 'a'. The argument goes as follows: if children have an innate syntactic determiner category, they should interchangeably use 'the' and 'a' with all nouns as soon as they begin producing them with a determiner. However, if children initially learn determiner-noun combinations as islands and only gradually abstract a syntactic category, they should initially use particular nouns with only one determiner (Valian, Solt, & Stewart, 2009). These two possibilities can be quantified as 'overlap': the number of nouns children produce with both 'a' and 'the', divided by the number of nouns children produce with either. If overlap is 0, children use each noun only with one of the two determiners, suggesting island-based learning. If overlap is 1, children use each noun with both determiners, suggesting a productive syntactic category. Results from this paradigm have been mixed. Some researchers find that children's overlap is low, suggesting that an abstract category of determiner is gradually constructed rather than being present from the start (Pine, Freudenthal, Krajewski, & Gobet, 2013). Others counter that children's overlap is not significantly different from their parents', suggesting an innate syntactic category (Valian et al., 2009).

Yang (2013) addresses an important problem with using overlap as a measure of productivity. As Valian et al. (2009) observe, the fewer times a noun appears, the more likely it will appear with only one determiner. Therefore, low overlap may simply be the consequence of many nouns appearing only few times.

Yang therefore uses the frequencies of noun types and determiners to predict expected overlap if determiners and nouns freely combine within these frequency constraints. His model accurately predicts empirical overlap values in early child language. Yang interprets this result as showing that from the start, children have an abstract determiner category. This finding has since been cited as evidence for innate syntactic categories (Bolhuis, Tattersall, Chomsky, & Berwick, 2014).

We replicate Yang's model on the six children from the CHILDES corpus analysed in Yang (2013). We show that while the model holds on average across nouns, it poorly predicts the behaviour of individual nouns. As a result, it systematically underestimates the overlap that would occur if nouns and determiners freely combined within Zipfian constraints. Keeping constant the overall frequencies of nouns and determiners, we shuffle each child's productions so that determiners and nouns combine at random. For these shuffled data, overlap measures exceed those predicted by Yang's model. The model, then, predicts the children's data not because they resemble the product of a freely combinatorial grammar, but because determiners and nouns do **not** freely combine: many mid- to high-frequency nouns appear with only one determiner. While Yang acknowledges these 'use asymmetries', he characterises them as 'unlikely to be linguistic'. We argue, however, that a) these asymmetries significantly constrain both children's and adults' data, and b) they are linguistic, specifically the product of lexical semantics interacting with the discourse functions of 'a' and 'the'. Since the target of acquisition is therefore not a freely combinatorial system, but one conditioned on semantics and discourse factors, children's productions are more accurately represented as a gradual acquisition of these factors, rather than as either islands or grammatical combinations isolated from discourse. More broadly, studies using naturalistic corpora to test hypotheses about language acquisition and evolution should be wary of either taking constrained usage patterns as evidence of lack of grammar, or abstracting away from them in aid of revealing underlying rules, since these constraints are a non-arbitrary part of the function of the language.

## References

- Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved? *PLoS Biology*, *12*(8), e1001934.
- Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, *127*(3), 345–360.
- Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language*, *36*(4), 743–778.
- Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6324–7.